The Fragment Constant Method for Predicting Octanol–Air Partition Coefficients of Persistent Organic Pollutants at Different Temperatures

Xuehua Li, Jingwen Chen,[a) Li Zhang, Xianliang Qiao, and Liping Huang

Department of Environmental Science and Technology, Dalian University of Technology, Linggong Road 2, Dalian 116024, People's Republic of China

The octanol–air partition coefficient (K_{OA}) is a key physicochemical parameter for describing the partition of organic pollutants between air and environmental organic phases. Experimental determination of K_{OA} is costly and time consuming, and sometimes restricted by lack of sufficiently pure chemicals. There is a need to develop a simple but accurate method to estimate K_{OA}. In the present study, a fragment constant model based on five fragment constants and one structural correction factor, was developed for predicting log K_{OA} at temperatures ranging from 10 to 40°C. The model was validated as successful by statistical analysis and external experimental log K_{OA} data. Compared to other quantitative structure–property relationship methods, the present model has the advantage that it is much easier to implement. As aromatic compounds that contain C, H, O, Cl, and Br atoms, were included in the training set used to develop the model, the current fragment model applies to a wide range of chlorinated and brominated aromatic pollutants, such as chlorobenzenes, polychlorinated naphthalenes, polychlorinated biphenyls, polychlorinated dibenzo-*p*-dioxins and dibenzofurans, polycyclic aromatic hydrocarbons, and polybrominated diphenyl ethers, all of which are typical persistent organic pollutants. Further study is necessary to expand the utility of the method to all halogenated aliphatic and aromatic compounds. © *2006 American Institute of Physics*. [DOI: 10.1063/1.2203356]

CONTENTS

[a)]Author to whom correspondence should be addressed; electronic mail: jwchen@dl.cn

List of Figures

1. Introduction

The octanol–air partition coefficient (K_{OA}), defined as the ratio of solute concentration in air versus octanol when the octanol–air system is at equilibrium, has been used extensively for describing the partitioning of organic compounds between air and terrestrial organic phases that may include organic carbon in soil,[1–3] the waxy cuticle and lipid portion of vegetation,[4–6] the organic film of atmosphere particulate matter,[7,8] and even indoor carpet.[9] K_{OA} has strong temperature dependence,[10] which can be described by

$$\log K_{OA} = A + \frac{\Delta H_{OA}}{2.303RT}, \tag{1}$$

where A is the intercept; ΔH_{OA} is the enthalpy change involved in octanol to air transfer of a chemical; R is the ideal gas constant, and T is absolute temperature. This temperature dependence is very important for assessing the potential long-range transport of persistent organic pollutants (POPs).[11] K_{OA} was shown to be a key physicochemical property pertinent to the long-term arctic contamination potential of POPs, and relatively volatile (log $K_{OA}<9$) and water soluble substances are subject to transport to the arctic regions.[12]

In 1995, Harner and Mackay[10] measured K_{OA} values of selected chlorobenzenes (CBs), polychlorinated biphenyls (PCBs), and p,p′-DDT by a newly developed generator column method. Using the same method K_{OA} values were later determined for more PCBs,[13] polycyclic aromatic hydrocarbons (PAHs) and polychlorinated naphthalenes (PCNs),[2] polychlorinated dibenzo-p-dioxin/dibenzofurans (PCDD/Fs),[14] polybrominated diphenyl ethers (PBDEs),[1] and organochlorine pesticides (OPs).[15] CBs, PCBs, PCDD/Fs, PBDEs, OPs, and many PAHs are typical POPs for which the environmental levels and behavior are research focus of scientists worldwide. Gas chromatographic (GC) retention,[16–19] fugacity meter methods,[20] solid-phase microextraction (SPME),[21] and head-space gas-chromatographic (HS-GC) measurements[22] were also developed for K_{OA} determination. However, these experimental methods usually need special equipment, sufficiently pure chemicals, a great deal of expendables, and time, which cannot meet the needs for environmental fate assessment of the ever-increasing number of POPs.

K_{OA} can also be estimated from the octanol–water partition coefficient (K_{OW}) and Henry's law constant (H). There is, however, a possible error inherent in this estimation in addition to the obvious combination of the measurement errors in K_{OW} and H.[1] This method is also restricted by the lack of K_{OW}, H, and their temperature dependence data for many organic pollutants.[1] It is thus preferable to determine or estimate K_{OA} directly. Chen *et al.*[23–26] developed a series of quantitative predictive models for estimating K_{OA} using theoretical molecular structural descriptors including quantum chemical descriptors. Nevertheless the predictive models look complex due to the quantum chemical computations and thus are not convenient for practical estimation.

An alternative approach for developing predictive models of K_{OA} is the fragment constant method, which is based on the assumption that a property of organic compounds is dependent on the presence of some fragments, each of them making a contribution into it.[27] According to Leo,[28] a fragment refers to an atom, or atoms, whose exterior bonds are to isolating carbon atoms, and an isolating carbon is one that either has four single bonds, at least two of which are to nonheteroatoms or is multiply bonded to other carbon atoms.

The only input necessary for this approach is the chemical structures. Furthermore, the method has good interpretability.[27] The fragment constant method has been successfully used to predict physicochemical properties including K_{OW},[29] organic carbon normalized sorption coefficients for soils or sediments,[30] bioconcentration factors,[31] median effective concentrations,[32] vapor pressure and activity coefficients in water and octanol,[33] boiling points,[27] and retention indices.[27] K_{OA} is a free-energy based parameter that should be dependent on the structure of a chemical in an additive-constitutive fashion.[34] The purpose of this study is to develop predictive models for K_{OA} of POPs such as chlorinated and brominated aromatic pollutants using the fragment constant method.

TABLE 1. Illustrations on how to partition molecular structures. The arrows indicate the "joint C atom"

Examples	Fragment Set (II)	Fragment Set (III)
	$2f_{C*}^{\phi}, 8f_{CH}^{\phi}$	$2f_{C*}^{\phi}, 8f_{CH}^{\phi}$
	$2f_{C*}^{\phi}, 10f_{CH}^{\phi}$	$2f_{C**}^{\phi}, 10f_{CH}^{\phi}$
	$4f_{C*}^{\phi}, 8f_{CH}^{\phi}, 1f_{O}^{\phi\phi}$	$2f_{C**}^{\phi}, 1f_{C*-O-C*}^{\phi\phi}, 8f_{CH}^{\phi}$
	$2f_{C*}^{\phi}, 10f_{CH}^{\phi}, 1f_{O}^{\phi\phi}$	$1f_{C*-O-C*}^{\phi\phi}, 10f_{CH}^{\phi}$

2. Experimental Techniques for K_{OA} Determination and the Available Data

The experimental methods used for the measurement of K_{OA} can be classified as direct and indirect. The generator column method, fugacity meter measurements, SPME and HS-GC measurements are direct methods for determination of K_{OA}. The generator column and fugacity meter methods are applicable to semivolatile compounds. So far most of the K_{OA} and its temperature-dependence data have been determined using the generator column method.[1,2,10,13–15] The fugacity meter measurement,[20] a method similar to the generator column method, was used to determine K_{OA} values of ten PCB congeners at 25 °C. The SPME method was once used to determine K_{OA} values for hydroxy alkyl nitrates, 1,2-dichlorobenzene and phenanthrene at 25 °C.[21] The HS-GC measurements tend to be limited to fairly volatile organic compounds, which were used for K_{OA} determination of 74 volatile hydrocarbons at 25 °C.[22]

The direct methods are time consuming, especially at low temperatures, and involve several analytical steps, such as the extraction of the traps, concentration of analytes, and quantification against a calibration curve, which have the potential to introduce errors to the measured K_{OA} value.[16] Thus the relative GC retention index method was developed to determine K_{OA} indirectly for semivolatile organic compounds. For example, Wania et al.[16] and Lei et al.[19] used GC retention time method to determine K_{OA} values for PCBs, PCNs, and PBDEs,[16] and polyfluorinated sulfonamide, sulfonamidoethanols, and telomere alcohols.[19] The prerequisites of these GC methods are the knowledge of the temperature-dependent K_{OA} of a standard reference compound and directly measured K_{OA} values at one temperature for a sufficient number of calibration compounds.[16] Thus the accuracy of the indirect K_{OA} determination method rests with the data quality of the reference or calibration compounds. Errors from the reference or calibration compounds may lead to systematic errors for the indirectly determined K_{OA} values.

3. Development of the Fragment Constant Method

3.1. Training and Validation Data Set

The training set was selected based on the following rules: The K_{OA} values were directly measured; the temperature dependence data for K_{OA} are available; and, only the halogenated aromatic compounds (persistent organic pollutants) are considered in the current study. To develop predictive models covering aliphatic compounds and especially halogenated aliphatic compounds that are of importance in environmental studies, more directly determined K_{OA} and its temperature dependence data, in addition to the 74 values for volatile hydrocarbons at 25 °C,[22] are required. As a result, only the K_{OA} values directly determined by the generator column method were selected in the training set. The training set includes 238 log K_{OA} values at four typical environmental temperatures (10, 20, 30, and 40 °C), corresponding to 72 compounds including CBs,[10,15] PCBs,[10,13] PAHs,[2] PCNs,[2] PCDD/Fs,[14] and PBDEs.[1] These data have been widely used in environmental behavior assessment of POPs.[4–9,35,36] As the experimental log K_{OA} values for four PBDE congeners, PBDE-153, PBDE-154, PBDE-156, and PBDE-183 were identified as outliers in previous studies,[23,37] they were not included in the training set.

The K_{OA} values determined by other methods except the generator column method were included in the validation set, which includes log K_{OA} values for ten distinct PCB congeners at 25 °C determined by Kömp and McLachlan[20] using fugacity meter measurements, for 104 PCBs at 20 °C determined by Zhang et al.[17] using a multicolumn method, for PCDD/Fs at 25 °C extrapolated (I) and determined semiempirically from retention indices (II) by Harner et al.,[14] for six CBs and 27 PCNs from 10 to 40 °C determined by Su et al.[18] using the isothermal capacity factors, and for selected PCBs, PCNs, and PBDEs at 25 °C determined by Wania et al.[16] employing the retention index method. Generally, these data are consistent with the corresponding values determined by the generator column method. For example, the difference between the PCB log K_{OA} values measured by the generator

TABLE 2. Illustrations on how to determine the number of occurrences (m_j) of structural correction factors

Structures	Number of occurrences for the specific structural correction factor (m_j)
	2,2′,4,4′,5-Pentachlorobiphenyl For $F_{2,6}$, $m_{2,6} = 2$. Positions: 2-, 2'-. For $F_{3,5}$, $m_{3,5} = 1$. Positions: 5-.
	1,2,3,7,8-PCDD For F_α, $m_\alpha = 1$. Position: 1-. For F_β, $m_\beta = 4$. Positions: 2-, 3-, 7-, 8-.

column method[13] and the fugacity meter method[20] averaged 0.3 log units.[16] Deviations between the log K_{OA} values of Wania *et al.* and those determined by the generator column method were on average 0.2 log units, and never larger than 0.55 log units.[16] In addition, statistically significant and precise correlations were reported between the log K_{OA} values determined by the generator column method and the GC retention indices, as indicated by the squared regression coefficient (r^2) and standard deviation (SD). The r^2 values reported by Zhang *et al.*[17] and Su *et al.*[18] were in the range 0.980–0.997, and SD is in the range 0.007–0.220.

3.2. Fragmentation Method

According to Leo,[28] a single-atom fragment can only be an isolating carbon atom or a hydrogen or heteroatom (e.g., –H, –O–). A multiple-atom fundamental fragment is any combination of nonisolating carbon, hydrogen, and/or heteroatoms (e.g., –CH, –C–O–C–).[28] It is essential to guarantee that the fragments of a chemical must not be selected arbitrarily. Herein the compounds under study are substituted aromatic hydrocarbons. Thus three sets of fragment constants were put forward and evaluated for their significance in the model. The first set consists of single-atom fragment constants, f_C^Φ, f_H^Φ, f_{Cl}^Φ, f_{Br}^Φ, and $f_O^{\Phi\Phi}$, which stand for the corresponding atoms in an aromatic ring or bond to an aromatic ring. The superscript Φ indicates that all the fragments are in or bond to an aromatic ring, and when it is used twice, bond to an aromatic ring on two sides. The second and third sets of fragments consist of multiple-atom fragments. The second set includes f_{CH}^Φ, f_{CCl}^Φ, f_{CBr}^Φ, f_{C*}^Φ, and $f_O^{\Phi\Phi}$, where f_{CH}^Φ, f_{CCl}^Φ, and f_{CBr}^Φ stand for two-atom fragments in an aromatic ring, f_{C*}^Φ represents the "joint C atom" defined as a single C atom in an aromatic ring that bonds to aromatic C or O atoms only, and $f_O^{\Phi\Phi}$ stands for a single O atom fragment bonding to two aromatic C atoms, as illustrated in Table 1. The third set includes fragment constants f_{CH}^Φ, f_{CCl}^Φ, f_{CBr}^Φ, f_{C**}^Φ, and $f_{C*-O-C*}^{\Phi\Phi}$, where the "joint C atom" in the second set was further classified as f_{C**}^Φ that represents the "joint C atom" fragment bonding to aromatic C atom only. If the "joint C atom" in the second set bonds to an O atom, it is merged into the three-atom fragment constant $f_{C*-O-C*}^{\Phi\Phi}$.

3.3. Structural Correction Factors

Besides the constitutional effects characterized by the fragments, the effects of steric features on K_{OA} should also be taken into consideration when characterizing compounds with a relatively complex structure. For POPs like PCBs, their properties and biological activities are quite different with respect to the coplanar or noncoplanar structures.[38–42] PCBs with chlorine atoms at the 2-, 2′-, 6- and 6′- positions are noncoplanar.[38–40,42] The extent of noncoplanarity can be described indirectly with planarization energy, which means the energy difference between the coplanar and minimum energy conformations.[41] The planarization energy for PCBs as well as PBDEs with two or three halogen atoms at the 2-, 2′-, 6-, and 6′- positions is higher than those with naught or one halogen atom.[41,42] The positioning of chlorine atoms in the 3- and/or 5- positions reduces the extent to which ortho-substituents can bend back at the equilibrium position, thus the dihedral angle between the rings is greater than if the 3- and 5- positions are not substituted.[43] So two structural correction factors, $F_{2,6}$ and $F_{3,5}$, were included to characterize the nonplanar steric effects. In addition, two structural correction factors, F_α and F_β, which denote the substituents at the ortho (α) or meta (β) positions, were screened to characterize the influence of halogen atoms for planar compounds like PCDD/Fs and PCNs. The number of occurrences (m_j) for the respective structural correction factor (j) is defined as the number of substituents at the specific positions. Examples are presented in Table 2 to illustrate how to count m_j of the structural correction factors.

3.4. Model Development

log K_{OA} of a compound with known structure can be calculated using the following equation:

$$\log K_{OA} = \sum_{i=1}^{a} n_i f_i + \sum_{j=1}^{b} m_j F_j, \qquad (2)$$

where a and b represent the total number of the fragments and structural correction factors, respectively; n_i and m_j are the number of occurrence for the ith fragment and the jth structural correction factor; f_i is the fragment constant for the ith fragment; and F_j is the structural factor value for the jth structural feature. For the training compounds, values of n_i and m_j were available, thus multiple regression (MLR) was employed to estimate the values of f_i and F_j, by evaluating and using the most significant regression equations.

In MLR, multicollinearity among the input variables may result in wrong signs and magnitudes of regression coefficient estimates that are the resulting f_i and F_j values in the current study. Thus the variance inflation factor (VIF), which measures how much the variance of the standardized regression coefficient is inflated by multicollinearity, was adopted to evaluate multicollinearity among the input variables. VIF for variable X_k is defined as

TABLE 3. Statistical parameters for the three sets of fragments obtained by MLR at 20 °C. N: number of log K_{OA} values in the training set. F: the statistic of F test. R^2_{adj}: coefficient of determination adjusted by degree of freedoms. As it was obtained by regression analysis about the origin, it cannot be compared to R^2 for models that include an intercept. SE: standard errors of the estimated values. VIF: variance inflation factor.

Fragment set (I)	f_C^Φ	f_H^Φ	f_{Cl}^Φ	f_{Br}^Φ	$f_O^{\Phi\Phi}$
t statistics	10.120	−3.359	4.168	10.020	0.644
Significance level (p)	<0.001	<0.001	<0.001	<0.001	0.523
VIF	262.05	96.98	48.92	4.43	3.09
$N=72$, $F=5584$ ($p<0.001$), $R^2_{adj}=0.997$, SE=0.476.					
Fragment set (II)	$f_{C^*}^\Phi$	f_{CH}^Φ	f_{CCl}^Φ	f_{CBr}^Φ	$f_O^{\Phi\Phi}$
t statistics	10.120	14.370	57.228	33.078	0.644
Significance level (p)	<0.001	<0.001	<0.001	<0.001	0.523
VIF	13.89	8.09	2.50	1.71	3.09
$N=72$, $F=5584$ ($p<0.001$), $R^2_{adj}=0.997$, SE=0.476.					
Fragment set (III)	$f_{C^{**}}^\Phi$	f_{CH}^Φ	f_{CCl}^Φ	f_{CBr}^Φ	$f_{C^*\text{-O-}C^*}^{\Phi\Phi}$
t statistics	10.120	14.370	57.228	33.078	15.606
Significance level (p)	<0.001	<0.001	<0.001	<0.001	<0.001
VIF	7.89	8.09	2.50	1.71	2.66
$N=72$, $F=5584$ ($p<0.001$), $R^2_{adj}=0.997$, SE=0.476.					

$$(VIF)_k = \frac{1}{1 - R_k^2}, \tag{3}$$

where R_k^2 stands for coefficient of determination for X_k when it is predicted by the other independent variables included in the MLR equation. VIF values exceeding 10 are often regarded as serious multicollinearity.[44]

The statistical significance of MLR models can be characterized by statistics such as the F statistic from the analysis of variance, standard errors (SE) of the estimated values, coefficient of determination adjusted by degree of freedoms (R^2_{adj}), and the significance levels (p) that represent the probability of error that is involved in accepting an observed result as valid. Thus the higher the F and R^2_{adj} values, the higher is the significance of a model; the lower the SE value, the greater is the precision of the model; and the lower the p value, the higher is the reliability of the model.

4. Results

4.1. The Significant Fragment Set

To evaluate the significance and goodness of fit for the three sets of fragments, MLR analysis was performed using n_i as independent variables only, at four typical environmental temperatures (10, 20, 30, 40 °C). For brevity, only the statistical results at 20 °C are listed Table 3, which shows the three sets of fragments resulted in similar overall statistics, such as F statistic, R^2_{adj}, and SE. However for the first and second fragment sets, $f_O^{\Phi\Phi}$ seems not statistically significant ($p=0.523$), which may be due to the lower occurrence number of $f_O^{\Phi\Phi}$ in the training molecules than other fragments. In addition, the VIF for f_C^Φ is as high as 262, and VIF for $f_{C^*}^\Phi$ is 13.9, indicating strong multicollinearity between the independent variables included in the two MLR equa-

tions, which can lead to incorrect estimates of the fragment constants or parameter estimates are artificially statistically nonsignificant. Thus the first and second sets of fragments were excluded from the subsequent discussions.

For the third fragment set, all the predictor variables are statistically significant ($p<0.001$), and all the VIF values are lower than 10, implying that the third fragment set is the best one and overcomes the problem of multicollinearity and thus the values of the fragment constants are genuine.

4.2. The Structural Correction Factors

After the optimal fragment set having been selected, the structural correction factors were evaluated for their necessity in the modeling of K_{OA}, using stepwise variable selection regression analysis. The t statistics for $F_{3,5}$, F_α, and F_β are 0.60 ($p=0.55$), 1.78 ($p=0.08$), and 0.22 ($p=0.83$), respectively, indicating that these structural correction factors are statistically insignificant. In the final regression model, all of the independent variables (five fragment constants of the third set and $F_{2,6}$) are significant at the $p<0.001$ level. Thus the third set of fragments together with $F_{2,6}$ is the best combination in explaining log K_{OA}. Similar statistical analysis was performed for the other temperatures and other possible combinations between the fragment sets and the structural corrections, which gave a similar conclusion.

4.3. The Final Fragment Constant Model

The final statistical results based on the third fragment set and $F_{2,6}$, for the four environmental temperatures, are summarized in Table 4, which shows that all the four regression results are statistically significant ($p<0.001$). Figure 1 shows scatter plots of the observed versus fitted values of log K_{OA}, which gives a visual impression of how strongly these two values are related. Quantitative assessment of the consistence can be described by the simple correlation coefficients (r) listed in Table 4. As shown in Fig. 1 and indicated by the high r (>0.983) values, the log K_{OA} values predicted by the fragment constant models are quite consistent with the observed ones, suggesting that the fragment constant method is successful in estimating log K_{OA}. The SE values range
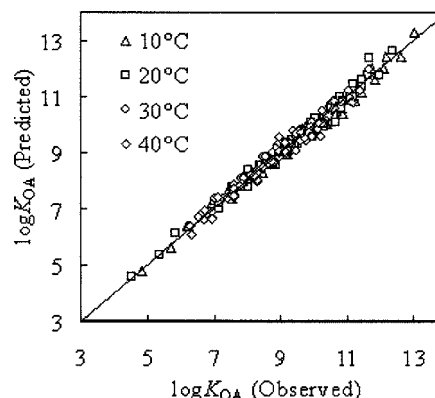

FIG. 1. Plot of predicted against observed log K_{OA} values at different temperatures.

TABLE 5. The f_i and F_j values and SE at four typical environmental temperatures. The number of occurrence for $f_{C^{**}}^{\Phi}$, f_{CH}^{Φ}, f_{CCl}^{Φ}, f_{CBr}^{Φ}, $f_{C^*\text{-}O\text{-}C^*}^{\Phi\Phi}$ and $F_{2,6}$ in the models are: 87, 292, 200, 38, 20, and 45 at 10 °C; 107, 331, 261, 38, 30, and 45 at 20 °C; 99, 282, 200, 38, 20, and 39 at 30 °C; 69, 218, 154, 38, 30, and 13, at 40 °C; respectively. The f_i and F_j values at 25 °C are calculated by the temperature dependence Eq. (4). The values in brackets are SE values.

Temperatures	$f_{C^{**}}^{\Phi}$ (SE)	f_{CH}^{Φ} (SE)	f_{CCl}^{Φ} (SE)	f_{CBr}^{Φ} (SE)	$f_{C^*\text{-}O\text{-}C^*}^{\Phi\Phi}$ (SE)	$F_{2,6}$ (SE)
10 °C	0.748 (0.042)	0.530 (0.016)	1.333 (0.012)	1.807 (0.024)	1.606 (0.065)	−0.435 (0.027)
20 °C	0.710 (0.038)	0.506 (0.015)	1.273 (0.011)	1.729 (0.024)	1.469 (0.054)	−0.435 (0.028)
30 °C	0.633 (0.049)	0.480 (0.020)	1.223 (0.013)	1.629 (0.025)	1.428 (0.080)	−0.398 (0.032)
40 °C	0.589 (0.052)	0.461 (0.022)	1.142 (0.013)	1.482 (0.034)	1.362 (0.071)	−0.213 (0.073)
25 °C	0.668	0.493	1.240	1.657	1.463	−0.368

from 0.207 to 0.223, which are considerably lower than in a previous study,[23] where SE=0.277 for a universal predictive model that included all the POPs under study and exploited many theoretical molecular descriptors as predictor variables.[23]

The resulting f_i and F_j values together with their SE values at the four typical environmental temperatures are listed in Table 5. The f_{CH}^{Φ} values are smaller than f_{CCl}^{Φ} and f_{CBr}^{Φ}, thus, with substitution of H atoms by Cl or Br atoms in a parent molecular structure, the log K_{OA} values increase. The $F_{2,6}$ values are negative, thus noncoplanar PCBs or PBDEs have much lower log K_{OA} values and tend to partition into the air phase.

4.4. Temperature Dependence for f_i and F_j

Given the temperature dependence of log K_{OA} expressed by Eq. (1), temperature dependence of f_i or F_j was investigated based on the following linear equation:
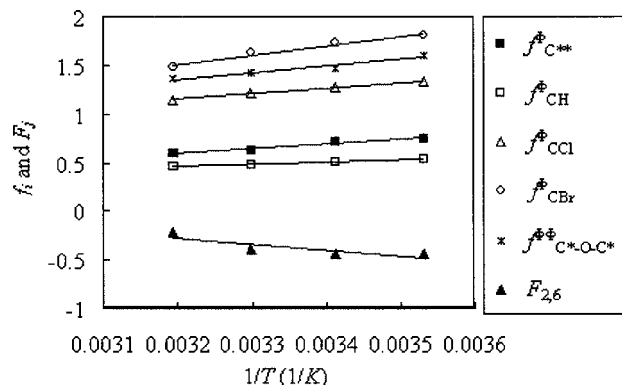
FIG. 2. The temperature dependences of f_i and F_j: ($f_{C^{**}}^{\Phi}$=490.49/T−0.977, r=0.990, $p<0.010$; f_{CH}^{Φ}=206.75/T−0.200, r=0.999, $p<0.001$; f_{CCl}^{Φ}=550.81/T−0.607, r=0.992, $p<0.008$; f_{CBr}^{Φ}=948.45/T−1.524, r=0.985, $p<0.016$; $f_{C^*\text{-}O\text{-}C}^{\Phi\Phi}$=689.24/$T$−0.849, r=0.974, $p<0.027$; $F_{2,6}$=−611.98/T +1.685, r=0.838, $p<0.162$).
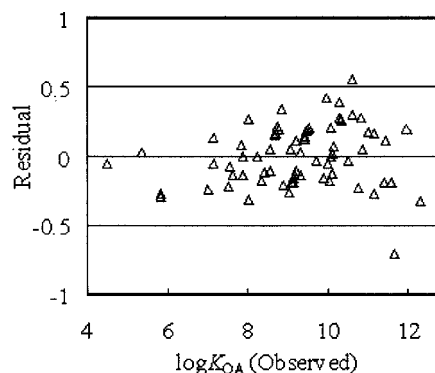
FIG. 3. Plot of the residuals against the training log K_{OA} values at 20 °C.

$$f_i \text{ (or } F_j) = s/T + q, \tag{4}$$

where s and q stand for regression parameters. The results are illustrated in Fig. 2. For $f_{C^*}^{\Phi}$, f_{CH}^{Φ}, f_{CCl}^{Φ}, f_{CBr}^{Φ}, and $f_{C^*\text{-}O\text{-}C^*}^{\Phi\Phi}$, significant ($p<0.05$) and strong (high s values) temperature dependences are observed. f_{CBr}^{Φ} has the highest temperature dependence among all the fragment constants. Generally f_i decrease with the increase of temperature and the converse is true for $F_{2,6}$. The fact that log K_{OA} values decrease with the increase of temperature[1,2,13,14] (i.e., in the same direction as f_i) indicates that the fragments play a larger role in governing the temperature dependence than the structural correction factor. Based on Eq. (4), f_i and F_j values at 25 °C were estimated, which are listed in Table 5 too.

5. Evaluations on the Final Fragment Constant Model

5.1. Residual Analysis for the Regression Models

The validity of the fragment constant models can be assessed by analysis of residuals. The residuals are the differences between the observed and predicted log K_{OA} values. The purpose of residual analysis is to test whether the residuals are randomly and normally distributed, and whether significant descriptor variables have been neglected from the models.[45]

Figure 3 shows the plot of residuals versus the training log K_{OA} values at 20 °C as an example. Inspection of the plot reveals that most of the data points (except for two

TABLE 4. The statistics of stepwise regression at four typical environmental temperatures. r: the simple correlation coefficient between the observed and fitted values. D-W statistics: Durbin–Watson test for a serial correlation (nonrandomness) of the residuals. D-W statistics between 1.5 and 2.5 indicate the residuals are independent.[44] The other statistics are the same as in Table 3.

Temperatures	N	F	R_{adj}^2	SE	r	D–W statistics
10 °C	60	21 430 ($p<0.001$)	0.999	0.207	0.993	2.257
20 °C	72	21 300 ($p<0.001$)	0.999	0.222	0.990	2.155
30 °C	58	15 665 ($p<0.001$)	0.999	0.223	0.983	2.263
40 °C	48	14 048 ($p<0.001$)	0.999	0.207	0.988	2.463
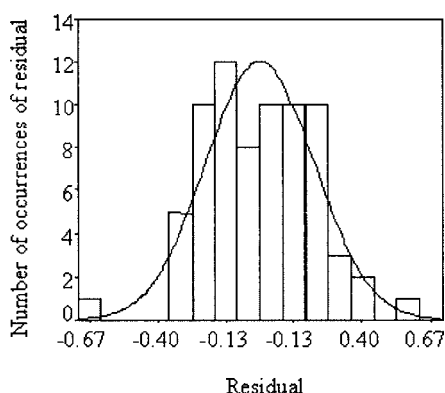
FIG. 4. Histogram of the number of occurrence of the residuals against the residuals for log K_{OA} at 20 °C.

points) lie between -0.5 and 0.5 log units and are randomly scattered about zero; there are no systematic trends in the residuals indicative of errors in the model or anomalous values due to individual outliers. The Durbin–Watson (D–W) statistics can be used to test serial correlations (nonrandomness) of the residuals.[44] One of the assumptions for regression analysis is that the residuals for consecutive observations are uncorrelated. The expected value of the Durbin–Watson statistic is 2. Values less than 2 indicate the possibility of positive autocorrelations; and values greater than 2 indicate negative autocorrelations. As a rule of thumb, D–W statistics between 1.5 and 2.5 indicate the values are independent.[44] The D–W statistics for serial correlations of the residuals are summarized in Table 4. All the D–W statistics are close to 2, indicating that the residuals for the consecutive observations are uncorrelated.

Although the lack of systematic trends in the residual plots suggests that the errors are randomly distributed, it does not demonstrate that the distribution is normal. This can be further verified by the histogram of residuals (Fig. 4), which plots the number of occurrence of the residuals versus the residuals. The bin width used to generate Fig. 4 was 0.090, and the total number of bins was 16. Figure 4 reveals a distinctive bell-shaped pattern associated with a normal distribution. Application of the Kolmogorov–Smirnov test for normality (at the 95% confidence level) confirms that the distribution shown in Fig. 4 is a normal distribution (mean $=0.000$, SD$=0.214$). Equivalent results were obtained for the residuals from the other environmental temperatures. The normal distribution of residuals implies that: (1) the residuals are nonsystematic, and (2) the fragment constants and structural correction factors are sufficient to explain the variance of log K_{OA} values, which assures the validity of the fragment constant models as well as the multiple regression analysis.

5.2. Validation of the Fragment Constant Model

A full list of the POPs under study, their experimental and predicted log K_{OA} values at different temperatures are given in Table 6. At least 78% of the compounds in the validation set were not included in the training set. Table 6 presents

results of the comparison and shows that the errors (log unit) are generally smaller than 0.25 for 75% of the compounds.

As shown by Fig. 5 that compares the experimental values in the validation set with those calculated by the fragment constant method, strong consistency exists between the two sets of values. Figure 6 plots the prediction errors defined as differences between observed and predicted log K_{OA} values, versus the observed log K_{OA} values in the validation set. Inspection of the plot reveals that more than 97% of the data points lie between -0.5 and 0.5 log units and are randomly scattered about zero. Mean absolute error and SD of the prediction errors are 0.180 and 0.221, respectively. The histogram of the predictive errors for the validation set was shown in Fig. 7, for which the bin width was 0.094, and the total number of bins was 19. Application of the Kolmogorov–Smirnov test for normality (at the 95% confidence level) confirms that the distribution is normal. A modified jackknife test[46] was also performed for the compounds under study, which showed a high degree of robustness of the fragment constant method too. Thus the fragment constant method can predict log K_{OA} at temperatures ranging from 10 to 40 °C with success. In view of the scarceness of chemical standards for some POPs, the difficulty in experimental determinations, and the high cost involved in experimental determinations, the fragment constant method could serve as a fast, simple, and prior approach for calculating log K_{OA} values.

6. Sample Calculations for K_{OA}

A few sample calculations based on seven representative compounds for which the corresponding number of occurrence for fragments and structural correction factors are listed in Table 7, are included to demonstrate how the chemicals were fragmented to derive the predicted results, as follows:

(a) log K_{OA} (for pentachlorobenzene at 10 °C) $=f_{CH}^{\Phi}$ $+5f_{CCl}^{\Phi}=0.530+5\times1.333=7.195$. The corresponding experimental value is 6.931.[10]

(b) log K_{OA} (for 1,4-dichloronaphthalene at 10 °C) $=2f_{C^{**}}^{\Phi}+6f_{CH}^{\Phi}+2f_{CCl}^{\Phi}=2\times0.748+6\times0.530+2$ $\times1.333=7.342$. The corresponding experimental value is 7.524.[2]

(c) log K_{OA} (for 2,2′,5,6′-tetrachlorobiphenyl at 30 °C) $=2f_{C^{**}}^{\Phi}+6f_{CH}^{\Phi}+4f_{CCl}^{\Phi}+3F_{2,6}=2\times0.633+6$ $\times0.480+4\times1.223-3\times0.398=7.844$. The corresponding experimental value is 7.84.[13]

(d) log K_{OA} (for 1,2,3,4,7-P$_5$CDD at 20 °C) $=3f_{CH}^{\Phi}$ $+5f_{CCl}^{\Phi}+2f_{C^*-O-C^*}^{\Phi\Phi}=3\times0.506+5\times1.273+2\times1.469$ $=10.821$. The corresponding experimental value is 10.996.[14]

(e) log K_{OA} (for 2,3,7,8-T$_4$CDF at 40 °C) $=2f_{C^{**}}^{\Phi}$ $+4f_{CH}^{\Phi}+4f_{CCl}^{\Phi}+f_{C^*-O-C^*}^{\Phi\Phi}=2\times0.589+4\times0.461+4$ $\times1.142+1.362=8.952$. The corresponding experimental value is 9.348.[14]

(f) log K_{OA} (for 2,2′,4,4′-BDE at 25 °C) $=6f_{CH}^{\Phi}$ $+4f_{CBr}^{\Phi}+f_{C^*-O-C^*}^{\Phi\Phi}+2F_{2,6}=7\times0.493+3\times1.657$

TABLE 6. The log K_{OA} values of selected POPs at different temperatures

Compounds	t (°C)	Observed	This study	Validation set		Error
1-Chloronaphthalene	10		6.539	6.39[a]		−0.15
	20		6.235	6.10[a]		−0.14
	25		6.027			
	30		5.849	5.52[a]		−0.33
	40		5.547	5.30[a]		−0.25
2-Chloronaphthalene	10		6.539	6.36[a]		−0.18
	20		6.235	6.08[a]		−0.16
	25		6.027			
	30		5.849	5.50[a]		−0.35
	40		5.547	5.28[a]		−0.27
1,2-Dichloronaphthalene	10		7.342	7.35[a]		0.01
	20		7.002	7.01[a]		0.01
	25		6.774		6.89[f]	0.12
	30		6.592	6.44[a]		−0.15
	40		6.228	6.13[a]		−0.10
1,4-Dichloronaphthalene	10	7.524	7.342			
	20	7.134	7.002			
	25		6.774		6.78[f]	0.01
	30	6.716	6.592			
	40	6.380	6.228			
1,5-Dichloronaphthalene	10		7.342	7.26[a]		−0.08
	20		7.002	6.92[a]		−0.08
	25		6.774			
	30		6.592	6.36[a]		−0.23
	40		6.228	6.06[a]		−0.17
2,7-Dichloronaphthalene	10		7.342	7.28[a]		−0.06
	20		7.002	6.95[a]		−0.05
	25		6.774			
	30		6.592	6.38[a]		−0.21
	40		6.228	6.08[a]		−0.15
1,2,3-Trichloronaphthalene	10		8.145	8.24[a]		0.10
	20		7.769	7.85[a]		0.08
	25		7.521			
	30		7.335	7.30[a]		−0.04
	40		6.909	6.91[a]		0.00
1,2,5-Trichloronaphthalene	10		8.145	8.12[a]		−0.03
	20		7.769	7.74[a]		−0.03
	25		7.521			
	30		7.335	7.19[a]		−0.15
	40		6.909	6.81[a]		−0.10
1,2,6-Trichloronaphthalene	10		8.145	8.16[a]		0.02
	20		7.769	7.77[a]		0.00
	25		7.521			
	30		7.335	7.22[a]		−0.12
	40		6.909	6.83[a]		−0.08
1,2,7-Trichloronaphthalene	10		8.145	8.19[a]		0.04
	20		7.769	7.80[a]		0.03
	25		7.521			
	30		7.335	7.25[a]		−0.09
	40		6.909	6.86[a]		−0.05
1,6,7-Trichloronaphthalene	10		8.145	8.19[a]		0.04
	20		7.769	7.80[a]		0.03
	25		7.521			
	30		7.335	7.25[a]		−0.09
	40		6.909	6.86[a]		−0.05
1,2,3,4-Tetrachloronaphthalene	10		8.948	9.03[a]		0.08

TABLE 6. The log K_{OA} values of selected POPs at different temperatures—Continued

Compounds	t (°C)	Observed	This study	Validation set		Error
	20		8.536	8.59[a]		0.05
	25		8.268		8.30[f]	0.03
	30		8.078	8.05[a]		−0.03
	40		7.590	7.59[a]		0.00
1,2,3,5-Tetrachloronaphthalene	10		8.948	8.98[a]		0.03
	20		8.536	8.55[a]		0.01
	25		8.268		8.29[f]	0.02
	30		8.078	8.00[a]		−0.08
	40		7.590	7.55[a]		−0.04
1,2,3,7-Tetrachloronaphthalene	10		8.948	9.05[a]		0.10
	20		8.536	8.62[a]		0.08
	25		8.268			
	30		8.078	8.07[a]		−0.01
	40		7.590	7.61[a]		0.02
1,2,3,8-Tetrachloronaphthalene	10		8.948	9.37[a]		0.42
	20		8.536	8.92		0.38
	25		8.268			
	30		8.078	8.38[a]		0.30
	40		7.590	7.89[a]		0.30
1,2,4,5-Tetrachloronaphthalene	10		8.948			
	20	8.867	8.536			
	25		8.268			
	30	8.294	8.078			
	40	7.881	7.590			
1,2,4,6-Tetrachloronaphthalene	10	8.788	8.948			
	20	8.360	8.536			
	25		8.268			
	30	7.818	8.078			
	40	7.393	7.590			
1,2,4,7-Tetrachloronaphthalene	10		8.948			
	20		8.536			
	25		8.268		8.13[f]	−0.14
	30		8.078			
	40		7.590			
1,2,4,8-Tetrachloronaphthalene	10	9.164	8.948			
	20	8.690	8.536			
	25		8.268			
	30	8.140	8.078			
	40	7.688	7.590			
1,2,5,6-Tetrachloronaphthalene	10		8.948	8.95[a]		0.00
	20		8.536	8.53[a]		−0.01
	25		8.268			
	30		8.078	7.98[a]		−0.10
	40		7.590	7.53[a]		−0.06
1,2,5,8-Tetrachloronaphthalene	10	9.140	8.948			
	20	8.699	8.536			
	25		8.268			
	30	8.143	8.078			
	40	7.685	7.590			
1,2,6,7-Tetrachloronaphthalene	10		8.948	9.05[a]		0.10
	20		8.536	8.62[a]		0.08
	25		8.268			
	30		8.078	8.07[a]		−0.01
	40		7.590	7.61[a]		0.02
1,2,7,8-Tetrachloronaphthalene	10		8.948	9.44[a]		0.49
	20		8.536	8.99[a]		0.45

LI *ET AL.*

TABLE 6. The log K_{OA} values of selected POPs at different temperatures—Continued

Compounds	t (°C)	Observed	This study	Validation set	Error
	25		8.268		
	30		8.078	8.45[a]	0.37
	40		7.590	7.95[a]	0.36
1,3,5,7-Tetrachloronaphthalene	10		8.948	8.58[a]	−0.37
	20		8.536	8.18[a]	−0.36
	25		8.268		
	30		8.078	7.62[a]	−0.46
	40		7.590	7.21[a]	−0.38
1,3,5,8-Tetrachloronaphthalene	10		8.948	8.98[a]	0.03
	20		8.536	8.55[a]	0.01
	25		8.268		
	30		8.078	8.00[a]	−0.08
	40		7.590	7.55[a]	−0.04
1,3,6,8-Tetrachloronaphthalene	10		8.948	8.95[a]	0.00
	20		8.536	8.53[a]	−0.01
	25		8.268		
	30		8.078	7.98[a]	−0.10
	40		7.590	7.53[a]	−0.06
1,4,5,8-Tetrachloronaphthalene	10	9.188	8.948		
	20	8.750	8.536		
	25		8.268		
	30	8.190	8.078		
	40	7.724	7.590		
1,4,6,7-Tetrachloronaphthalene	10	8.845	8.948		
	20	8.418	8.536		
	25		8.268		
	30	7.868	8.078		
	40	7.430	7.590		
2,3,6,7-Tetrachloronaphthalene	10		8.948	9.16[a]	0.21
	20		8.536	8.72[a]	0.18
	25		8.268		
	30		8.078	8.17[a]	0.09
	40		7.590	7.70[a]	0.11
1,2,3,4,5-Pentachloronaphthalene	10		9.751	10.07[a]	0.32
	20		9.303	9.58[a]	0.28
	25		9.015		
	30		8.821	9.05[a]	0.23
	40		8.271	8.49[a]	0.22
1,2,3,4,6-Pentachloronaphthalene	10	9.744	9.751		
	20	9.201	9.303		
	25		9.015	8.92[f]	−0.10
	30	8.625	8.821		
	40	8.117	8.271		
1,2,3,5,6-Pentachloronaphthalene	10	10.079	9.751		
	20	9.480	9.303		
	25		9.015		
	30	8.804	8.821		
	40	8.301	8.271		
1,2,3,5,7-Pentachloronaphthalene	10	9.502	9.751		
	20	9.041	9.303		
	25		9.015	8.82[f]	−0.20
	30	8.465	8.821		
	40	7.968	8.271		
1,2,3,5,8-Pentachloronaphthalene	10	9.973	9.751		
	20	9.438	9.303		
	25		9.015	9.10[f]	0.08

TABLE 6. The log K_{OA} values of selected POPs at different temperatures—Continued

Compounds	t (°C)	log K_{OA}			
		Observed	This study	Validation set	Error
	30	8.857	8.821		
	40	8.338	8.271		
1,2,3,6,7-Pentachloronaphthalene	10		9.751	9.87[a]	0.12
	20		9.303	9.39[a]	0.09
	25		9.015		
	30		8.821	8.85[a]	0.03
	40		8.271	8.32[a]	0.05
1,2,3,6,8-Pentachloronaphthalene	10		9.751	9.98[a]	0.23
	20		9.303	9.50[a]	0.20
	25		9.015		
	30		8.821	8.96[a]	0.14
	40		8.271	8.40[a]	0.13
1,2,4,5,7-Pentachloronaphthalene	10	9.633	9.751		
	20	9.167	9.303		
	25		9.015		
	30	8.593	8.821		
	40	8.104	8.271		
1,2,4,5,8-Pentachloronaphthalene	10		9.751		
	20	9.493	9.303		
	25		9.015		
	30	8.891	8.821		
	40	8.352	8.271		
1,2,4,6,7-Pentachloronaphthalene	10		9.751		
	20		9.303		
	25		9.015	9.58[f]	0.57
	30		8.821		
	40		8.271		
1,2,4,6,8-Pentachloronaphthalene	10	9.537	9.751		
	20	9.111	9.303		
	25		9.015		
	30	8.522	8.821		
	40	8.004	8.271		
1,2,4,7,8-Pentachloronaphthalene	10	9.807	9.751		
	20	9.423	9.303		
	25		9.015		
	30	8.831	8.821		
	40	8.274	8.271		
1,2,3,4,5,6-Hexachloronaphthalene	10		10.554		
	20	10.326	10.070		
	25		9.762		
	30	9.945	9.564		
	40	9.346	8.952		
1,2,3,4,5,7-Hexachloronaphthalene	10		10.554		
	20	10.072	10.070		
	25		9.762		
	30	9.572	9.564		
	40	8.950	8.952		
1,2,3,4,5,8-Hexachloronaphthalene	10		10.554		
	20	10.622	10.070		
	25		9.762		
	30	10.207	9.564		
	40	9.559	8.952		
1,2,3,4,6,7-Hexachloronaphthalene	10	10.576	10.554		
	20	10.013	10.070		
	25		9.762		
	30	9.462	9.564		

TABLE 6. The log K_{OA} values of selected POPs at different temperatures—Continued

| Compounds | t (°C) | log K_{OA} | | | |
		Observed	This study	Validation set	Error
	40	8.838	8.952		
1,2,3,5,6,7-Hexachloronaphthalene	25		9.762	9.58[f]	−0.18
1,2,3,5,7,8-Hexachloronaphthalene	10		10.554		
	20	10.090	10.070		
	25		9.762	9.67[f]	−0.09
	30	9.616	9.564		
	40	8.988	8.952		
1,2,4,5,6,8-Hexachloronaphthalene	10		10.554		
	20	10.140	10.070		
	25		9.762	9.69[f]	−0.07
	30	9.670	9.564		
	40	9.053	8.952		
1,2,3,4,5,6,7-Heptachloronaphthalene	10		11.357	11.52[a]	0.16
	20		10.837	10.96[a]	0.12
	25		10.509	10.38[f]	−0.13
	30		10.307	10.44[a]	0.13
	40		9.633	9.75[a]	0.12
1,2,3,4,5,6,8-Heptachloronaphthalene	10		11.357	11.56[a]	0.20
	20		10.837	10.99[a]	0.15
	25		10.509		
	30		10.307	10.47[a]	0.16
	40		9.633	9.79[a]	0.16
1,2,3,4,5,6,7,8-Octachloronaphthalene	10		12.160	12.39[a]	0.23
	20		11.604	11.78[a]	0.18
	25		11.256	11.05[f]	−0.21
	30		11.050	11.27[a]	0.22
	40		10.314	10.51[a]	0.20
Chlorobenzene	10		3.983	3.76[a]	−0.22
	20		3.803	3.45[a]	−0.35
	25		3.705		
	30		3.623	3.17[a]	−0.45
	40		3.447	2.90[a]	−0.55
1,2-Dichlorobenzene	10	4.820	4.786		
	20	4.510	4.570		
	25		4.452		
	30		4.366		
	40		4.128		
1,3-Dichlorobenzene	10		4.786	4.60[a]	−0.19
	20		4.570	4.27[a]	−0.30
	25		4.452		
	30		4.366	3.96[a]	−0.41
	40		4.128	3.67[a]	−0.46
1,4-Dichlorobenzene	10		4.786	4.65[a]	−0.14
	20		4.570	4.32[a]	−0.25
	25		4.452		
	30		4.366	4.01[a]	−0.36
	40		4.128	3.72[a]	−0.41
1,2,3-Trichlorobenzene	10	5.699	5.589		
	20	5.365	5.337		
1,2,4-Trichlorobenzene	10		5.589	5.45[a]	−0.14
	20		5.337	5.10[a]	−0.24
	25		5.199		
	30		5.109	4.77[a]	−0.34
	40		4.809	4.46[a]	−0.35
1,3,5-Trichlorobenzene	10		5.589	5.23[a]	−0.36
	20		5.337	4.89[a]	−0.45

TABLE 6. The log K_{OA} values of selected POPs at different temperatures—Continued

| Compounds | t (°C) | log K_{OA} | | | |
		Observed	This study	Validation set	Error
	25		5.199		
	30		5.109	4.56[a]	−0.55
	40		4.809	4.26[a]	−0.55
1,2,3,4-Tetrachlorobenzene	10	6.213	6.392		
	20	5.818	6.104		
1,2,3,5-Tetrachlorobenzene	10		6.392	6.15[a]	−0.24
	20		6.104	5.78[a]	−0.32
	25		5.946		
	30		5.852	5.43[a]	−0.42
	40		5.490	5.11[a]	−0.38
1,2,4,5-Tetrachlorobenzene	10	6.204	6.392		
	20	5.829	6.104		
Pentachlorobenzene	10	6.931	7.195		
	18.7	6.539			
PCB-0	20		6.480	6.09[b]	−0.39
PCB-1	20		6.812	6.65[b]	−0.16
PCB-2	20		7.247	7.00[b]	−0.25
PCB-3	30	6.62	6.809		
	25		7.013	6.80[f]	−0.21
	20	7.01	7.247	6.99[b]	−0.26
	10	7.43	7.599		
PCB-4	20		7.144	6.86[b]	−0.28
PCB-5	20		7.579	7.59[b]	0.01
PCB-6	20		7.579	7.55[b]	−0.03
PCB-7	20		7.579	7.39[b]	−0.19
PCB-8	20		7.579	7.61[b]	0.03
PCB-9	20		7.579	7.40[b]	−0.18
PCB-11	20		8.014	7.90[b]	−0.11
PCB-12	20		8.014	7.80[b]	−0.21
PCB-14	20		8.014	7.78[b]	−0.23
PCB-15	25		7.760	7.73[f]	−0.03
	20	7.88	8.014	7.88[b]	−0.13
	10	8.35	8.402		
PCB-16	20		7.911	7.98[b]	0.07
PCB-17	20		7.911	7.74[b]	−0.17
PCB-18	25		7.771	7.60[c]	−0.17
	20		7.911	7.79[b]	−0.12
PCB-20	20		8.346	8.49[b]	0.14
PCB-22	20		8.346	8.58[b]	0.23
PCB-25	20		8.346	8.28[b]	−0.07
PCB-26	20		8.346	8.27[b]	−0.08
PCB-28	20		8.346	8.40[b]	0.05
PCB-29	25		8.139	8.01[f]	−0.13
	20	8.03	8.346	8.05[b]	−0.30
	10	8.51	8.770		
PCB-31	20		8.346	8.40[b]	0.05
PCB-32	20		7.911	7.97[b]	0.06
PCB-33	20		8.346	8.52[b]	0.17
PCB-41	20		8.678	8.82[b]	0.14
PCB-44	25		8.518	8.36[c]	−0.16
	20		8.678	8.71[b]	0.03
PCB-46	20		8.243	8.56[b]	0.32
PCB-47	20		8.678	8.56[b]	−0.12
PCB-48	20		8.678	8.50[b]	−0.18
PCB-49	30	8.21	8.242		
	20	8.57	8.678	8.63[b]	−0.05

LI *ET AL.*

TABLE 6. The log K_{OA} values of selected POPs at different temperatures—Continued

Compounds	t (°C)	Observed	This study	Validation set		Error
	10	9.08	9.138			
PCB-52	25		8.518	8.22[c]	8.47[f]0.30(−0.05)	
	20		8.678	8.49[b]		−0.19
PCB-53	30	7.84	7.844			
	20	8.24	8.243	8.18[b]		−0.06
	10	8.70	8.703			
PCB-61	25		8.886		8.80[f]	−0.09
	20	8.90	9.113	8.93[b]		−0.18
	10	9.40	9.573			
PCB-63	20		9.113	9.06[b]		−0.05
	10		9.573			
PCB-64	25		8.518	8.41[b]		−0.11
	20		8.678	8.63[b]		−0.05
	10		9.138			
PCB-66	30	8.82	8.640			
	20	9.22	9.113	9.29[b]		0.18
	10	9.65	9.573			
PCB-70	20		9.113	9.22[b]		0.11
PCB-71	20		8.678	8.84[b]		0.16
PCB-74	20		9.113	9.14[b]		0.03
PCB-77	30	9.47	9.038			
	25		9.254		9.29[b]	0.04
	20	9.96	9.548	9.92[b]		0.37
	10	10.36	10.008			
PCB-83	20		9.445	9.39[b]		−0.05
PCB-84	20		9.010	9.28[b]		0.27
PCB-95	30	8.55	8.587			
	25		8.897	8.71[c]		−0.19
	20	9.06	9.010	9.06[b]		0.05
	10	9.51	9.506			
PCB-96	30	8.30	8.189			
	20	8.77	8.575	8.79[b]		0.22
	10	9.22	9.071			
PCB-97	20		9.445	9.44[b]		−0.01
PCB-101	30	8.78	8.985			
	25		9.265		9.14[f]	−0.13
	20	9.31	9.445	9.28[b]		−0.17
	10	9.78	9.941			
PCB-105	30	9.77	9.383			
	25		9.633			
	20	10.27	9.880	10.20[b]		0.32
	10	10.84	10.376			
PCB-110	25		9.265	9.06[c]		−0.21
	20		9.445	9.58[b]		0.14
PCB-118	30	9.57	9.383			
	20	10.08	9.880	10.04[b]		0.16
	10	10.64	10.376			
PCB-126	30	10.10	9.781			
	20	10.61	10.315	10.66[b]		0.35
	10	11.24	10.811			
PCB-131	20		9.777	9.83[b]		0.05
PCB-132	20		9.777	10.07[b]		0.29
PCB-134	20		9.777	9.71[b]		−0.07
PCB-135	20		9.777	9.69[b]		−0.09
PCB-136	20		9.342	9.53[b]		0.19
PCB-138	30	9.53	9.728			

TABLE 6. The log K_{OA} values of selected POPs at different temperatures—Continued

Compounds	t (°C)	Observed	This study	Validation set		Error
	20	10.09	10.212	10.20[b]		−0.01
	10	10.61	10.744			
PCB-141	20		10.212	10.07[b]		−0.14
PCB-144	20		9.777	9.62[b]		−0.16
PCB-146	20		10.212	9.84[b]		−0.37
PCB-147	20		9.777	9.67[b]		−0.11
PCB-149	25		9.644	9.27[c]		−0.37
	20		9.777	9.74[b]		−0.04
PCB-151	20		9.777	9.58[b]		−0.20
PCB-153	30	9.39	9.728			
	25		10.012	9.37[c]	9.80[f]0.64(−0.21)	
	20	10.04	10.212	9.99[b]		−0.22
	10	10.62	10.744			
PCB-155	20	9.16	9.342	9.13[b]		−0.21
	10	9.64	9.874			
PCB-156	20		10.647	10.87[b]		0.22
PCB-157	20		10.647	11.07[b]		0.42
PCB-158	20		10.212	10.14[b]		−0.07
PCB-163	20		10.212	10.16[b]		−0.05
PCB-167	20		10.647	10.77[b]		0.12
PCB-169	20		11.082	11.32[b]		0.24
PCB-170	20		10.979	11.07[b]		0.09
PCB-171	30	9.96	10.073			
	20	10.51	10.544	10.51[b]		−0.03
	10	11.14	11.112			
PCB-172	20		10.979	10.67[b]		−0.31
PCB-173	20		10.544	10.60[b]		0.06
PCB-174	20		10.544	10.51[b]		−0.03
PCB-175	20		10.544	10.17[b]		−0.37
PCB-176	20		10.109	10.06[b]		−0.05
PCB-177	20		10.544	10.58[b]		0.04
PCB-178	20		10.544	10.12[b]		−0.42
PCB-179	20		10.109	10.10[b]		−0.01
PCB-180	30	10.23	10.471			
	25		10.759	9.88[c]		−0.88
	20	10.75	10.979	10.72[b]		−0.26
	10	11.38	11.547			
PCB-183	20		10.544	10.26[b]		−0.28
PCB-187	25		10.391	9.87[c]		−0.52
	20		10.544	10.22[b]		−0.32
PCB-189	20		11.414	11.54[b]		0.13
PCB-190	20		10.979	10.87[b]		−0.11
PCB-191	20		10.979	10.91[b]		−0.07
PCB-193	20		10.979	10.82[b]		−0.16
PCB-194	20		11.746	11.59[b]		−0.16
PCB-195	20		11.311	11.44[b]		0.13
PCB-196	20		11.311	11.03[b]		−0.28
PCB-197	20		10.876	10.52[b]		−0.36
PCB-198	20		11.311	11.05[b]		−0.26
PCB-199	20		11.311	11.05[b]		−0.26
PCB-200	20		10.876	10.82[b]		−0.06
PCB-201	20		10.876	10.98[b]		0.10
PCB-202	20		10.876	10.38[b]		−0.50
PCB-203	20		11.311	11.10[b]		−0.21
PCB-205	20		11.746	11.62[b]		−0.13
PCB-206	20		12.078	11.79[b]		−0.29

TABLE 6. The log K_{OA} values of selected POPs at different temperatures—Continued

Compounds	t (°C)	Observed	This study	Validation set		Error
				log K_{OA}		
PCB-207	20		11.643	11.26[b]		−0.38
PCB-208	20		11.643	11.26[b]		−0.38
PCB-209	20		12.410	11.96[b]		−0.45
1-CDD	10	8.466	8.255			
	20	8.018	7.753			
	25		7.617	7.86[d]		0.24
	30	7.629	7.439			
	40	7.396	7.093			
2,3-D2CDD	25		8.364		8.50[e]	0.14
2,7-D2CDD	10	9.020	9.058			
	20	8.564	8.520			
	25		8.364	8.36[d]	8.48[e]	−0.00(0.12)
	30	8.106	8.182			
	40	7.818	7.774			
2,8-D2CDD	10	9.020	9.058			
	20	8.564	8.520			
	25		8.364	8.36[d]	8.48[e]	−0.00(0.12)
	30	8.106	8.182			
	40	7.818	7.774			
1,2,4-T3CDD	25		9.111		8.97[e]	−0.14
2,3,7-T3CDD	10	9.816	9.861			
	20	9.313	9.287			
	25		9.111	9.14[d]	9.42[e]	0.03(0.31)
	30	8.935	8.925			
	40	8.497	8.455			
1,2,3,4-T4CDD	10	10.400	10.664			
	20	9.896	10.054			
	25		9.858	9.70[d]	9.64[e]	0.16(−0.22)
1,2,3,7-T4CDD	25		9.858		9.94[e]	0.08
1,3,6,8-T4CDD	25		9.858		9.38[e]	−0.48
2,3,7,8-T4CDD	20	10.318	10.054			
	25		9.858	10.05[d]	9.95[e]	0.19(0.09)
	40	9.283	9.136			
1,2,3,4,7-P5CDD	20	10.996	10.821			
	25		10.605	10.67[d]	10.42[e]	0.07(−0.19)
	40	9.751	9.817			
1,2,3,7,8-P5CDD	20	10.867	10.821			
	25		10.605	10.57[d]	10.46[e]	0.04(−0.15)
	40	9.755	9.817			
1,2,3,4,7,8-H6CDD	20	11.403	11.588			
	25		11.352	11.11[d]	10.95[e]	0.24(−0.40)
	40	10.297	10.498			
1,2,3,6,7,8-H6CDD	25		11.352		10.97[e]	−0.38
1,2,3,7,8,9-H6CDD	25		11.352		11.01[e]	−0.34
1,2,3,4,6,7,8-H7CDD	10		13.073			
	20	11.660	12.355			
	25		12.099	11.42[d]	11.45[e]	0.68(−0.65)
	30		11.897			
	40	10.774	11.179			
2,3,7,8-T4CDF	10	10.829	10.554			
	20	10.281	10.005			
	25		9.731	10.02[d]	9.82[e]	0.29(0.09)
	30	9.707	9.506			
	40	9.348	8.952			
2-PBDEs	25		7.189		7.24[f]	0.05

TABLE 6. The log K_{OA} values of selected POPs at different temperatures—Continued

Compounds	t (°C)	Observed	This study	Validation set	Error
3-	25		7.557	7.36[f]	−0.20
2,4-	25		8.353	8.37[f]	0.02
2,4′-	25		8.353	8.47[f]	0.12
2,6-	25		7.985	8.12[f]	0.14
3,4-	25		8.721	8.55[f]	−0.17
3,4′-	25		8.721	8.57[f]	−0.15
4,4′-	25		8.721	8.64[f]	−0.08
2,2′,4-	10	9.981	9.867		
	20	9.523	9.328		
	30	9.095	8.879		
	40	8.694	8.609		
2,3,4-	25		9.517	9.49[f]	−0.03
2,4,4′-	10	10.195	10.302		
	20	9.726	9.763		
	30	9.289	9.277		
	40	8.879	8.822		
2,4,6-	25		9.149	9.02[f]	−0.13
2,4′,6-	25		9.149	9.28[f]	0.13
3,3′,4-	25		9.885	9.61[f]	−0.28
3,4,4′-	25		9.885	9.68[f]	−0.21
2,2′,4,4′-	10	11.429	11.144		
	20	10.818	10.551		
	30	10.248	10.028		
	40	9.714	9.630		
	25	10.499	10.313	10.34[f]	0.03
2,3,4,4′-	25		10.681	10.49[f]	−0.19
2,3′,4,4′-	10	11.813	11.579		
	20	11.141	10.986		
	30	10.514	10.426		
	40	9.926	9.843		
	25	10.773	10.681		
2,3′,4,6-	25		10.313	10.23[f]	−0.08
2,4,4′,6-	25		10.313	10.13[f]	−0.18
3,3′,4,4′-	10	11.742	12.014		
	20	11.148	11.421		
	30	10.592	10.824		
	40	10.072	10.056		
	25	10.829	11.049	10.7[f]	−0.35
2,2′,3,3′,4-	25		11.477	11.14[f]	−0.34
2,2′,3,4,4′-	10	12.607	12.421		
	20	11.965	11.774		
	30	11.365	11.177		
	40	10.804	10.651		
2,2′,4,4′,5-	10	12.160	12.421		
	20	11.587	11.774		
	30	11.052	11.177		
	40	10.551	10.651		
	25	11.321	11.477	11.28[f]	−0.20
2,2′,4,4′,6-	10	12.100	11.986		
	20	11.442	11.339		
	30	10.828	10.779		
	40	10.253	10.438		
	25	11.185	11.109	11.52[f]	0.41
3,3′,4,4′,5-	10	13.052	13.291		
	20	12.320	12.644		
	30	11.636	11.973		

TABLE 6. The log K_{OA} values of selected POPs at different temperatures—Continued

| Compounds | t (°C) | log K_{OA} | | | |
		Observed	This study	Validation set	Error
	40	10.996	11.077		
2,2′,4,4′,5,5′-	10	12.731[g]	13.698		
	20	12.113[g]	12.997		
	30	11.536[g]	12.326		
	40	10.995[g]	11.672		
	25	11.860[g]	12.641	12.15[f]	
2,2′,4,4′,5,6′-	10	12.795[g]	13.263		
	20	12.201[g]	12.562		
	30	11.646[g]	11.928		
	40	11.126[g]	11.459		
2,3,3′,4,4′,5-	10	12.911[g]	14.133		
	20	12.273[g]	13.432		
	30	11.676[g]	12.724		
	40	11.118[g]	11.885		
2,2′,3,4,4′,5′,6-	10	12.79[g]	14.540		
	20	12.227[g]	13.785		
	30	11.701[g]	13.077		
	40	11.209[g]	12.480		
HCB	10	7.887	7.998		
	15	7.700			
	20	7.563	7.638		
	25	7.388	7.440		
fluorene	10	7.501	7.544		
	20	7.130	7.190		
	25		6.934		
	30	6.516	6.699		
	40	6.093	6.377		
phenanthrene	10	8.267	8.292		
	20	7.897	7.900		
	25		7.602		
	30	7.418	7.332		
	40	6.926	6.966		
pyrene	10	9.528	9.788		
	20	9.155	9.320		
	25		8.938		
	30	8.547	8.598		
	40	8.121	8.144		
fluoranthene	20	9.124	9.320		
	25		8.938		
	30	8.652	8.598		
	40	8.161	8.144		

Note: Observed: the log K_{OA} values determined by the generator column method.[1,2,10,13–15] This study: the log K_{OA} values calculated by the fragment constant method. Validation set: the log K_{OA} values used to verify the models. The error is the difference between observed and predicted log K_{OA} values in validation set.

[a]The log K_{OA} values determined by Su _et al._[18]
[b]The log K_{OA} values determined by Zhang _et al._[17]
[c]The log K_{OA} values presented by Kömp and McLachlan.[20]
[d]The log K_{OA} values extrapolated by Harner _et al._ (I).[14]
[e]The log K_{OA} values determined semiempirically by Harner _et al._(II).[14]
[f]The log K_{OA} values determined by Wania _et al._[16]
[g]The values were not included in the training set.

$+1.463-2\times0.368=10.313$. The corresponding experimental value is 10.499.[1]

(g) log K_{OA} (for fluorene at 40 °C) $=3f_{C^{**}}^{\Phi}+10f_{CH}^{\Phi}=3\times0.589+10\times0.461=6.377$. The corresponding experimental value is 6.093.[2]

It is also possible to estimate log K_{OA} values based on compounds with known log K_{OA} values. Here are two examples:
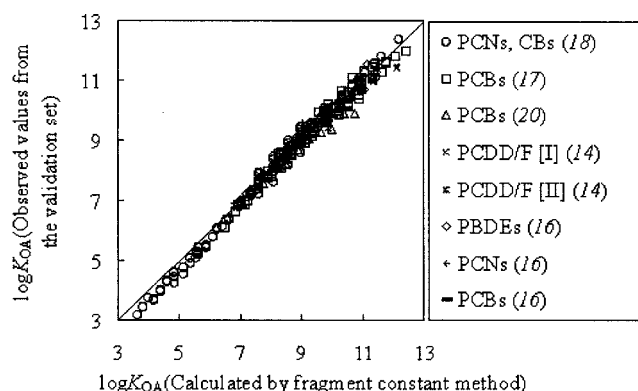
FIG. 5. Plot of log K_{OA} values calculated by the fragment constant method against observed values from the validation set.

(h) It is known that the experimental log K_{OA} value for 2,3,7,8-T$_4$CDD at 20 °C is 10.318.[14] The log K_{OA} of 1,2,3,7,8-P$_5$CDD at 20 °C can then be estimated as:

$$\log K_{OA}(1,2,3,7,8\text{-}P_5CDD \text{ at } 20 \text{ } ^\circ C)$$

$$= \log K_{OA}(2,3,7,8\text{-}T_4CDD \text{ at } 20 \text{ } ^\circ C) + f_{CCl}^{\Phi}$$

$$- f_{CH}^{\Phi} = 10.318 + 1.273 - 0.506 = 11.058.$$

The corresponding experimental value is 10.867.[14]

(i) It is known that the experimental log K_{OA} value for
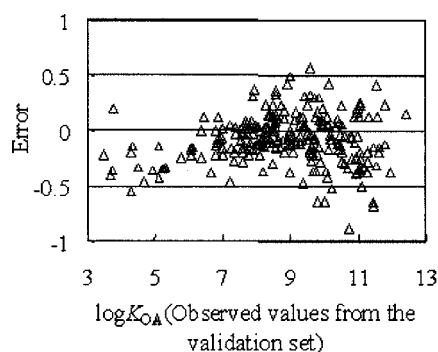
FIG. 6. Plot of the prediction errors against log K_{OA} values from validation set.
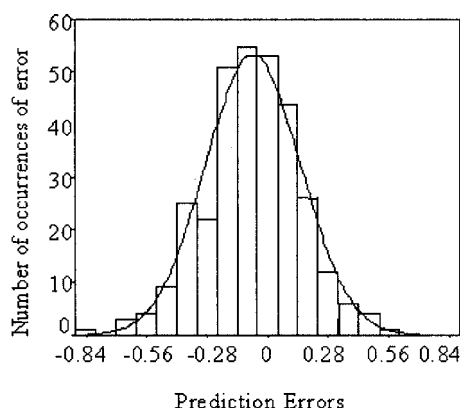
FIG. 7. Histogram of the number of occurrence of errors against the prediction errors of log K_{OA} from the validation set.

TABLE 7. Samples: Number of f_i or F_j for selected compounds

Compounds	Number of occurrence for f_i and F_j					
	$f_{C^{**}}^{\Phi}$	f_{CH}^{Φ}	f_{CCl}^{Φ}	f_{CBr}^{Φ}	$f_{C^*\text{-}O\text{-}C^*}^{\Phi\Phi}$	$F_{2,6}$
Pentachlorobenzene	0	1	5	0	0	0
1,4-Dichloronaphthalene	2	6	2	0	0	0
2,2′,5,6′-Tetrachlorobiphenyl	2	6	4	0	0	3
1,2,3,4,7-P$_5$CDD	0	3	5	0	2	0
2,3,7,8-T$_4$CDF	2	4	4	0	1	0
2,2′,4,4′-BDE	0	6	0	4	1	2
Fluorene	3	10	0	0	0	0

PCB-105 (2,3,3′,4,4′,-Pentachlorobiphenyl) at 30 °C is 9.77.[13] The log K_{OA} of PCB-77 (3,3′, 4,4′-Tetrachlorobiphenyl) at 30 °C can then be estimated as:

$$\log K_{OA}(\text{PCB-77 at } 30 \text{ } ^\circ C) = \log K_{OA}(\text{PCB}$$

$$\text{-105 at } 30 \text{ } ^\circ C) - f_{CBr}^{\Phi} + f_{CH}^{\Phi} - F_{2,6} = 9.77$$

$$- 1.629 + 0.480 + 0.398 = 9.02.$$

The corresponding experimental value is 9.47.[13]

7. Conclusion

In summary, a fragment constant model was developed for predicting log K_{OA} values at different environmental temperatures from 10 to 40 °C, which requires information on molecular structures only. Compared with other quantitative structure–property relationship (QSPR) models,[23] the current model has superior predictive ability and is much simpler to use. Thus the fragment constant model is ideal for predicting log K_{OA} for new aromatic compounds for which only limited data (such as molecular structures) is available. The current method can be used to predict log K_{OA} for chlorinated and brominated aromatic compounds, such as CBs, PCBs, PCDD/Fs, PCNs, PBDEs, and PAHs at different environmental temperatures. It can be inferred from the residual analysis and the external validation that the predicted values may have an error of ±0.5 log unit. As only aromatic compounds were involved in the training set, the current fragment model cannot be used for prediction of aliphatic compounds that have complex steric structures such as hexachlorocyclohexanes and heptachlor. Further study is necessary to expand the utility of the method to all halogenated aliphatic and aromatic compounds. For this purpose, more experimental K_{OA} data are necessary.

8. Acknowledgments

sincere gratitude to the anonymous reviewers and Dr. Chase for useful comments and kind attention.

9. References

[1] T. Harner and M. Shoeib, J. Chem. Eng. Data **47**, 228 (2002).

[2] T. Harner and T. F. Bidleman, J. Chem. Eng. Data **43**, 40 (1998).

[3] T. Harner, T. F. Bidleman, L. M. M. Jantunen, and D. Mackay, Envir. Toxicol. Chem. **20**, 1612 (2001).

[4] J. Tolls and M. S. McLachlan, Environ. Sci. Technol. **28**, 159 (1994).

[5] P. Kömp and M. S. McLachlan, Environ. Sci. Technol. **31**, 2944 (1997).

[6] J. A. Platts and M. H. Abraham, Environ. Sci. Technol. **34**, 318 (2000).

[7] D. Mackay, *Multimedia Environmental Models, The Fugacity Approach*, 2nd ed. (Lewis, 2001), p. 88.

[8] R. L. Falconer and T. Harner, Atmos. Environ. **34**, 4043 (2000).

[9] D. Won, R. L. Corsi, and M. Rynes, Environ. Sci. Technol. **34**, 4193 (2000).

[10] T. Harner and D. Mackay, Environ. Sci. Technol. **29**, 1599 (1995).

[11] C. Halsall, A. Sweetman, L. Barrie, and K. C. Jones, Atmos. Environ. **35**, 255 (2001).

[12] F. Wania, Environ. Sci. Technol. **37**, 1344 (2003).

[13] T. Harner and T. F. Bidleman, J. Chem. Eng. Data **41**, 895 (1996).

[14] T. Harner, N. J. L. Green, and K. Jones, Environ. Sci. Technol. **34**, 3109 (2000).

[15] M. Shoeib and T. Harner, Envir. Toxicol. Chem. **21**, 984 (2002).

[16] F. Wania, Y. D. Lei, and T. Harner, Anal. Chem. **74**, 3476 (2002).

[17] X. Zhang, K. W. Schramm, B. Henkelmann, C. Klimm, A. Kaune, A. Kettrup, and P. Lu, Anal. Chem. **71**, 3834 (1999).

[18] Y. Su, Y. D. Lei, G. L. Daly, and F. Wania, J. Chem. Eng. Data **47**, 449 (2002).

[19] D. Y. Lei, F. Wania, D. Mathers, and S. A. Mabury, J. Chem. Eng. Data **49**, 1013 (2004).

[20] P. Kömp and M. S. McLachlan, Envir. Toxicol. Chem. **17**, 2433 (1997).

[21] K. Treves, L. Shragina, and Y. Rudich, Atmos. Environ. **35**, 5843 (2001).

[22] H. M. Abraham, J. Le, W. E. Acree, Jr., P. W. Carr, and A. J. Dallas, Chemosphere **44**, 855 (2001).

[23] J. W. Chen, T. Harner, G. H. Ding, X. Quan, K-W. Schramm, and A. Kettrup, Envir. Toxicol. Chem. **23**, 2309 (2004).

[24] J. W. Chen, T. Harner, K-W. Schramm, X. Quan, X. Y. Xue, W. Z. Wu, and A. Kettrup, Comput. Biol. Chem. **27**, 405 (2003).

[25] J. W. Chen, T. Harner, K-W. Schramm, X. Quan, X. Y. Xue, W. Z. Wu, and A. Kettrup, Sci. Total Environ., **300**, 155 (2002).

[26] J. W. Chen, X. Y. Xue, K-W. Schramm, X. Quan, F. L. Yang, and A. Kettrup, Chemosphere **48**, 535 (2002).

[27] N. S. Zefirov and D. E. Palyulin, J. Chem. Inf. Comput. Sci. **42**, 1112 (2002).

[28] A. J. Leo, *Proceedings Symposium on Structure-Activity Correlations in Studies of Toxicity and Bio-Concentration with Aquatic Organisms*, Great Lakes Research Advisory Board, Burlington, Ontarion, 1975, p. 151.

[29] R. P. Schwarzenbach, P. M. Gschwend, and D. M. Imboden, *Environmental Organic Chemistry*, 2nd ed. (Wiley, New Jersey, 2003), pp. 228–235.

[30] S. Tao, H. S. Piao, R. Dawson, X. X. Lu, and H. Y. Hu, Environ. Sci. Technol. **33**, 2719 (1999).

[31] S. Tao, H. Y. Hu, X. X. Lu, R. W. Dawson, and F. L. Xu, Chemosphere **41**, 1563 (2000).

[32] S. Tao, X. H. Xi, F. L. Xu, B. G. Li, J. Gao, and R. Dawson, Environ. Pollut., **116**, 57 (2002).

[33] S. C. Lee, H. Hung, W-Y. Shiu, and D. Mackay, Envir. Toxicol. Chem. **19**, 2623 (2000).

[34] A. Leo, C. Hansch, and D. Elkins, Chem. Rev. (Washington, D.C.) **71**, 525 (1971).

[35] M. R. Chao, C. W. Hu, Y. L. Chen, G. P. Chang-Chien, W. J. Lee, L. W. Chang, W. S. Lee, and K. Y. Wu, Atmos. Environ. **38**, 1501 (2004).

[36] L. Nizzetto, C. Cassani, and A. Di Guardo, Ecotoxicol. Environ. Saf. **63**, 75 (2006).

[37] J. W. Chen, T. Harner, P. Yang, X. Quan, S. Chen, K-W. Schramm, and A. Kettrup, Chemosphere **51**, 577 (2003).

[38] M. Van den Berg, L. Birnbaum, A. T. C. Bosveld, B. Brunstrom, M. Feeley, R. Hasegawa, T. Kubiak, A. K. D. Liem, C. Nolt, F. X. R. Van Leeuwen, J. P. Giesy, S. W. Kennedy, P. Cook, A. Hanberg, and J. C. Larsen, Environ. Health Perspect. **106**, 775 (1998).

[39] M. T. O. Jonker and A. A. Koelmans, Environ. Sci. Technol. **36**, 3725 (2002).

[40] M. T. O. Jonker and F. Smedes, Environ. Sci. Technol. **34**, 1620 (2000).

[41] G. S. Chen, A. D. Konstantinov, B. G. Chittim, E. M. Joyce, N. C. Bols, and N. J. Bunce, Environ. Sci. Technol. **35**, 3749 (2001).

[42] P. R. Kodavanti, T. R. Ward, J. D. McKinney, C. L. Waller, and H. A. Tilson, Toxicol. Appl. Pharmacol. **138**, 251 (1996).

[43] G. R. Shaw and D. W. Connell, Environ. Sci. Technol. **18**, 18 (1984).

[44] M. J. Norusis, *SPSS 7.5 Guide to Data Analysis* (Simon and Schuster, Upper Saddle River, N.J., 1997), pp. 434–458.

[45] M. M. Scherer, B. B. Balko, D. A. Gallagher, and P. G. Tratnyek, Environ. Sci. Technol. **32**, 3026 (1998).

[46] M. T. Saçan, S. S. Erdem, G. A. Özpinar, and I. A. Balcioglu, J. Chem. Inf. Comput. Sci. **44**, 985 (2004).